Contract No. MDA 903-92-C-0126
Item No. 0002AL, Second Year Interim Technical Report

DRAFT

Technical Report-632

Structuring and Training High-Reliability Teams Year 2 Interim Technical Report

March 1994

Prepared by:

Jean MacMillan
Elliot E. Entin
Eileen B. Entin
Daniel Serfaty

ALPHATECH, Inc.
50 Mall Road
Burlington, MA 01803-4562

The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as official Agency position, policy, or decision, unless so designated by other official documentation.

Prepared for:

Dr. Dennis K. Leedom
U.S. Army Research Institute/Aviation R&D Activity/PERI-IR
Fort Rucker, AL 36362

DISTRIBUTION: Approved for public release; distribution is unlimited.

OFFICE OF THE UNDER SECRETARY OF DEFENSE (ACQUISITION & TECHNOLOGY)
DEFENSE TECHNICAL INFORMATION CENTER
8725 JOHN J KINGMAN RD STE 0944
FT BELVOIR VA 22060-6218

IN REPLY
REFER TO

DTIC-OMI

SUBJECT: Distribution Statements on Technical Documents

TO:

USARIA R&DA/PERI-IR
ATTN: DR. DENNIS K. LEEDOM
FORT RUCKER, AL 36362

1. Reference: DoD Directive 5230.24, Distribution Statements on Technical Documents, 18 Mar 87.

2. The Defense Technical Information Center received the enclosed report (referenced below) which is not marked in accordance with the above reference.

DRAFT TR632
MAR 94

3. We request the appropriate distribution statement be assigned and the report returned to DTIC within 5 working days.

4. Approved distribution statements are listed on the reverse of this letter. If you have any questions regarding these statements, call DTIC's Input Support Branch, (703) 767-9092, 9088 or 9086 (DSN use prefix 427).

FOR THE ADMINISTRATOR:

1 Encl

CRYSTAL RILEY
Chief, Input Support Branch

FL-171
Dec 95

DoD Directive 5230.24, "Distribution Statements on Technical Documents," 18 Mar 87, contains seven distribution statements, as described briefly below. Technical Documents that are sent to DTIC **must** be assigned one of the following distribution statements:

☒ **DISTRIBUTION STATEMENT A:**

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

☐ **DISTRIBUTION STATEMENT B:**

DISTRIBUTION AUTHORIZED TO U. S. GOVERNMENT AGENCIES ONLY; (FILL IN REASON); (DATE STATEMENT APPLIED). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (INSERT CONTROLLING DoD OFFICE).

☐ **DISTRIBUTION STATEMENT C:**

DISTRIBUTION AUTHORIZED TO U. S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (FILL IN REASON); (DATE STATEMENT APPLIED). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (INSERT CONTROLLING DoD OFFICE).

☐ **DISTRIBUTION STATEMENT D:**

DISTRIBUTION AUTHORIZED TO DoD AND DoD CONTRACTORS ONLY; (FILL IN REASON); (DATE STATEMENT APPLIED). OTHER REQUESTS SHALL BE REFERRED TO (INSERT CONTROLLING DoD OFFICE).

☐ **DISTRIBUTION STATEMENT E:**

DISTRIBUTION AUTHORIZED TO DoD COMPONENTS ONLY; (FILL IN REASON); (DATE STATEMENT APPLIED). OTHER REQUESTS SHALL BE REFERRED TO (INSERT CONTROLLING DoD OFFICE).

☐ **DISTRIBUTION STATEMENT F:**

FURTHER DISSEMINATION ONLY AS DIRECTED BY (INSERT CONTROLLING DoD OFFICE AND DATE), OR HIGHER DoD AUTHORITY.

☐ **DISTRIBUTION STATEMENT X:**

DISTRIBUTION AUTHORIZED TO U. S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DoD DIRECTIVE 5230.25 (DATE STATEMENT APPLIED). CONTROLLING DoD OFFICE IS (INSERT).

(Reason)	(Controlling DoD Office Name)
(Assigning Office)	(Controlling DoD Office Address (City/State/Zip)
(Signature & Typed Name)	(Date Statement Assigned)

STRUCTURING AND TRAINING HIGH-RELIABILITY TEAMS: YEAR 2 INTERIM TECHNICAL REPORT

EXECUTIVE SUMMARY

Research Requirement

The performance of teams working together to control complex systems has become increasingly important for a variety of military and civilian applications, and there is growing interest in identifying and understanding the factors that affect the reliability of such teams. A better understanding is needed of the coordination strategies used by superior teams to maintain acceptable performance in demanding environments in order to develop methods for structuring and training reliable teams.

Procedure

Data collected from experiments in a helicopter flight simulator were used to test a theory of team coordination for two-person helicopter flightcrews. The theory suggests that well-coördinated crews have congruent mental models of their mission, the situation, and each other. These congruent mental models support the crew's ability to communicate and coordinate effectively, leading to superior teamwork. This superior teamwork allows the crew to control their workload and keep it within manageable levels, leading to superior crew performance, especially in situations where external task demands are high. Experiment data are available on the congruence of the crew's mental models (from crew questionnaires), the quality of the crew's teamwork (rated by instructor pilots observing the crews), the subjective workload experienced by the crew (from crew rating forms), and the crew's performance (as evaluated by the instructor pilots). We tested the theory by examining the relationships among these variables.

Findings

All of the relationships predicted by the theory were found to be present in the data. Crews with more-congruent mental models exhibited superior teamwork, which was associated with lower perceived workloads. Superior teamwork and lower workloads were associated with higher levels of mission performance. In high-demand situations, crews exhibiting superior teamwork applied their effort more effectively, achieving superior mission performance at the same or lower workload levels than crews exhibiting lower levels of teamwork.

Utilization of Findings

The findings will be integrated with the results of an analysis of the crews' communication patterns to identify the communication and coordination strategies that are most effective for maintaining crew performance in high-demand situations. Strategies shown to be effective can be included in flightcrew training.

DRAFT

STRUCTURING AND TRAINING HIGH-RELIABILITY TEAMS:
YEAR 2 INTERIM TECHNICAL REPORT

CONTENTS

CONTENTS

LIST OF TABLES

CONTENTS

STRUCTURING AND TRAINING HIGH-RELIABILITY TEAMS: YEAR 2 INTERIM TECHNICAL REPORT

Introduction

ALPHATECH, Inc., under contract to the Army Research Institute Aviation Research & Development Activity (ARIARDA) is conducting research on structuring and training high-reliability teams. This interim report documents the initial activities performed during the second year of this project. These activities included the analysis of a series of measures based on a theory of team coordination. Data were collected and analyzed as part of two ARIARDA-sponsored activities—a Battle-Rostering Experiment conducted in October-November of 1993 and a preparatory pilot experiment conducted in June 1993.

Background

The performance of teams of operators controlling complex systems has become the focus of a growing research effort in recent years. There is increasing interest in identifying and understanding the factors that affect team performance and team errors in both military and civilian applications. The goal of this ARIARDA-sponsored research on structuring and training high-reliability teams is to provide insight into the development of such teams by examining how teams can best be structured and trained to support the flexible, adaptive behavior that has been observed to produce highly reliable team performance in real-world environments (LaPorte & Consolini, 1988; Pfeiffer, 1989; Reason, 1990). The effort has focused on the role of team-coordination strategies in producing reliable team performance.

Second Year Interim Activities

The Battle-Rostering Experiment provided us with an opportunity to test the team-coordination theory developed during the first year of the project and to apply the measurement instruments associated with that theory. The major activities for the second year of the project have been to test the measurement instruments and approaches during the pilot experiment, revise them as required, apply them in the Battle-Rostering Experiment, and analyze the results. This report presents the results of our analysis and assesses the implications of the findings for our theory of team coordination and team performance.

Organization of this Report

The remainder of this report summarizes our theory of crew coordination, describes the measurement methods used to collect data during the Battle-Rostering Experiment and its associated pilot experiment in order to test this theory, and presents our results. Our theoretical framework is used to organize the presentation of the experiment results. We test, in sequence, a

series of predictions about the relationships between the congruence of a crew's mental models, the quality of the crew's teamwork, the crew's perceived workload, and the crew's performance. The report concludes with a summary of the relationships found and a discussion of future directions for the research.

Theoretical Framework for Crew Coordination

A theoretical framework or model is needed to guide the selection and development of crew-coordination measures. Without such a model, there is no guidance as to what is important to measure in a crew-coordination study. The lack of theories to guide measurement is recognized as a general problem for team-performance research (Baker & Salas, 1992). One of the major activities during the first year of this project was the development of a theoretical framework for crew coordination and the specification of measures based on that framework (Entin, Entin, MacMillan, & Serfaty, 1993). This section briefly reviews the theoretical framework developed in Year 1.

One cannot understand the subtleties and richness of crew-coordination strategies without establishing links between crew-coordination strategies and crew performance, as well as identifying coordination strategies that are likely to lead to errors. Furthermore, since most critical crew errors seem to happen in periods of high workload, it is also important to understand the coping mechanisms that crews use to adapt to workload while attempting to maintain their effectiveness.

In order to meet these goals we need a theoretical construct that will link workload, crew processes (teamwork and taskwork), and performance (outcome). Figure 1 describes such a model proposed and validated by Serfaty and his colleagues (see, for example, Serfaty, Entin, & Volpe, 1993a, 1993b). It is based on the premise of *adaptation*: Superior crews cope with increases in workload through internal mechanisms of decision-strategy and coordination-strategy adaptation in an effort to keep team performance at the required level while maintaining workload at an acceptable level.

In the absence of a general theory of team behavior and crew coordination, a theoretical framework such as the one shown in Figure 1 can be used to link crew processes to team performance. The dynamic processes occurring in the cockpit during abnormal or crisis-like situations generate substantial levels of workload for the crew members. As a result, their behaviors and cognitive strategies—both individual and team-related—are strongly contingent upon the task environment. A good adaptation in coordination strategies may result in superior performance. On the other hand, a maladaptation or a lack of adaptability on the part of the crew may result in catastrophic errors.

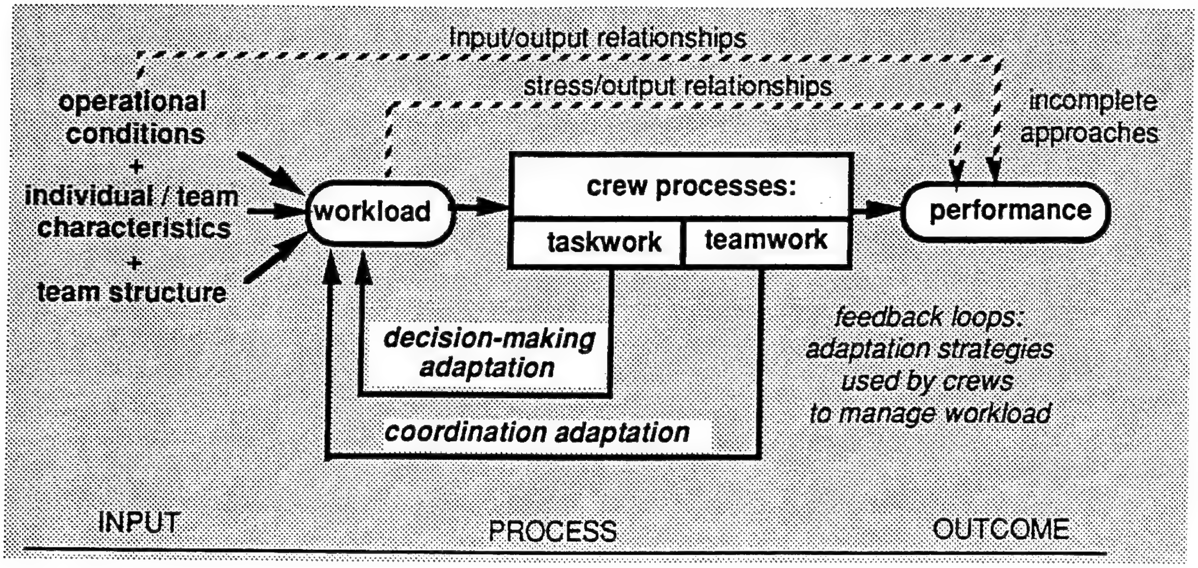


Figure 1. Theoretical framework for crew-coordination and crew-process dimensions and measures.

Adaptive team performance may be explained by an underlying theoretical premise: effective crews develop a mental model of their common task that enables them to use the team structure to maintain team coordination and performance under a wide range of conditions. It has been suggested by various authors (Cannon-Bowers, Salas, & Converse, 1990; Kleinman & Serfaty, 1989; MacIntyre, Morgan, Salas, & Glickman, 1988; Orasanu, 1990) that highly effective teams have a *shared mental model* of the situation and the task environment (consistent with the team's understanding of the situation), and *mutual mental models* of interacting team members' tasks and abilities that generate expectations about how other team members will behave. These mental models are particularly useful in the absence (or scarcity) of timely, error-free, and unambiguous information. We hypothesize that crews that have developed a high level of congruence between their mental models—both situational and mutual—are able to make use of these models to anticipate the way the situation will evolve as well as the needs of the other team members. These crews will perform consistently better under a wide range of flight conditions.

The coordination mechanisms that support adaptation may be *explicit*, based on specific communications, or *implicit*, based on shared or mutual mental models. Both explicit and implicit coordination will generate observable communication patterns—the presence *and* the absence of communication may be important. For example, communications that provide information to a team member in the absence of requests for that information indicate an implicit-coordination mechanism at work. Measures must be sensitive to changes in the team's coordination and communication patterns as they adapt their behavior to the demands of the task

and the environment. We expect to see teams shift between explicit coordination (under low-workload conditions) and implicit coordination (under high workload). Even though a team may have shared or mutual mental models that support implicit coordination, some explicit exchange of information will be required in order to maintain those models as the situation changes (Orasanu & Fischer, 1992). Observation has also suggested that members of highly reliable teams are aware of the workload of other team members and voluntarily assume some of the tasks of any individual in the team who is overloaded under stressful conditions. This dynamic reallocation of workload should be observable from the team's communication patterns.

Plan for Testing the Theory in Year 2

Figure 2 shows the methods and measures recommended in the Year 1 Technical Report for obtaining data to test the theoretical framework for flightcrew coordination. We recommended the use of existing crew-workload measurement approaches and the use of the previously developed performance measures for flightcrews based on the Aircrew Training Manual (ATM) as well as the use of mission-specific performance indices. For coordination and teamwork process measurement, we recommended both the use of existing teamwork-process measures such as the Aircrew Evaluation Checklist (ACE), and the development of new quantitative communications-analysis measures. During Year 1 we developed a communications-analysis data-collection instrument and methodology to provide these measures, and were able to test the instrument using videotapes of flightcrews during simulated flight (Entin, Entin, MacMillan, & Serfaty, 1993).

MEASURE / METHOD	WORKLOAD	CREW PROCESSES		PERFORMANCE (OUTCOME)
		TASKWORK	TEAMWORK	
SELF-REPORT (CREWS)	TLX		CMAQ	
AUTOMATED (SIMULATOR)		MISSION-SPECIFIC INDICES		MISSION-SPECIFIC INDICES
ON-SITE/OFF-SITE OBSERVATIONS (OBSERVERS)			COMMUNICATION ANALYSIS	
ON-SITE OBSERVATIONS (DOMAIN EXPERTS)		ACE EVALUATION QUESTIONNAIRES		ATM, SUBJ. EVAL.

Figure 2. Recommended measures for flightcrew-coordination research.

During Year 1 we also developed questionnaires to explicitly test the mental-model constructs underlying our theory of team coordination. In Year 2 we used these questionnaires in conjunction with an experiment sponsored by ARI on the effects of battle rostering for crews. The premise of the research is that *congruent mental models among crew members lead to superior crew performance*. This "congruence" has at least three dimensions:

- Congruence with "truth": i.e., how close to the truth are the two crew members' assessments of the situation?

- Congruence of the two situational mental models: i.e., how close to each other are the two crew members' assessments of the situation?

- Congruence of the mutual mental models: i.e., how well does one crew member anticipate/predict/understand the actions and information needs of the other?

Deficiencies in any of these three dimensions may be evident before, during, or after the mission. Therefore we collected data and assessed mental-model congruence at several points in time.

An associated hypothesis for the research is that battle rostering for crews may produce an exaggerated belief among crew members that they can anticipate and predict the actions and needs of the other crew members. This may lead to a variety of team errors that differ from the errors made by a crew that is not battle rostered.

The goals of our research effort were to assess the extent to which crew members held congruent mental models of the mission and of each other and to assess whether these congruent models led to higher performance. Figure 3 illustrates our hypotheses about the mechanisms by which congruent mental models improve crew performance. Our theory suggests that congruent mental models support the crew's ability to communicate and coordinate effectively, leading to superior teamwork. This superior teamwork allows the crew to control their workload and keep it within manageable levels, leading to superior crew performance, especially in situations where external task demands are high.

Another goal of our research was to assess the effects of battle rostering and crew-coordination training on crew behavior. Battle rostering and crew-coordination training may affect the crew's performance at one or more of the points shown in Figure 2. Crew-coordination training and/or a history of experience together as a battle-rostered team may increase the congruence of the crew's mental models. Alternatively, training or experience may improve the quality of the crew's teamwork and the level of the crew's performance without improving mental-model congruence.

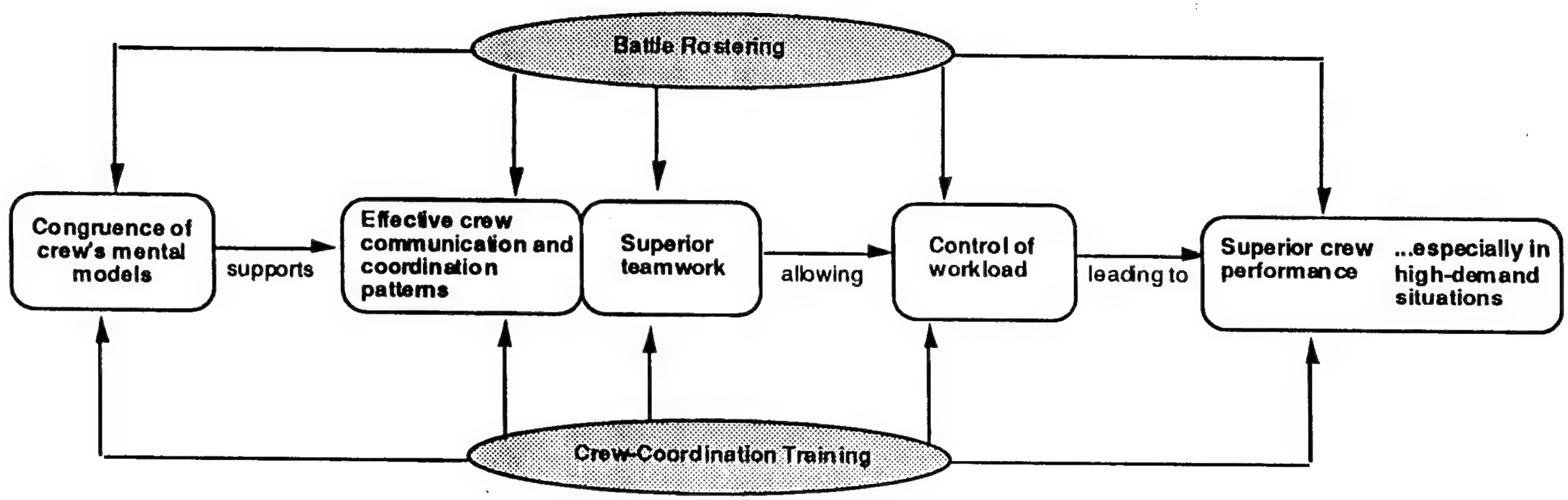


Figure 3. Hypothesized relationships between congruence of crew's mental models and crew performance, showing possible effects of battle rostering and crew-coordination training.

Data Collection and Measures

We had two opportunities to collect and analyze data during Year 2: a Battle-Rostering Experiment conducted to assess the performance of battle-rostered crews compared with the performance of mixed crews that were not battle rostered, and an earlier pilot experiment conducted to test the procedures and data collection plans for the full-scale experiment. This section describes the data available from these two efforts, including the data available from three questionnaires that were designed to collect data on the congruence of the crew's mental models and the aspects of teamwork expected to be most sensitive to that congruence. Our goal was to obtain measures that could be used to test each of the relationships shown in Figure 3.

Design of the Battle-Rostering Experiment

The purpose of the Battle-Rostering Experiment was to identify the relative contribution of the Army's battle-rostering policy, which keeps crews intact over time, and standardized crew-coordination training to the goals of mission performance and flight safety. The experiment was designed and conducted by Dynamics Research Corporation (DRC) in cooperation with ARIARDA, and is described in more detail in Grubb, Simon, Leedom, and Zeller (1994). Twelve two-person AH-64 Apache attack helicopter crews participated in the experiment. Each individual flew four missions in a flight simulator: two with his[1] usual partner (battle-rostered crew), and two with a different partner assigned from another crew (mixed crew). Four comparable scenarios were developed for the experiment (see Grubb, Simon, Leedom, & Zeller, 1994). The scenarios were designed to include several high-task-demand events, including equipment malfunctions as well as engagements with enemy forces. All of the crews participating in the experiment had previously received crew-coordination training (see Simon and Grubb (1993) for a description of the crew-coordination training program).

A pilot test was conducted four months prior to the Battle-Rostering Experiment in order to evaluate the experiment procedures. Eight battle-rostered AH-64 crews participated in the pilot experiment. These crews also flew four missions, but in the pilot experiment two of these missions were flown before the crews received crew-coordination training and two were flown after the training. Thus the pilot experiment provides data on the effects of the crew-coordination training for battle-rostered crews.

Measures of Mental-Model Congruence

We developed two questionnaires to assess crew members' mental-model congruence before and after they flew each mission.

[1]All of the aviators in the experiment were male.

The questionnaires were first used in the pilot experiment, and, based on the results of this pre-test, were slightly modified for use in the Battle-Rostering Experiment. Appendix A contains the final version of the questionnaires.

The Crew Member Pre-Mission Questionnaire was completed independently by each crew member prior to each scenario. Three items elicit elements of the crew members' mental models of critical aspects of the mission and their mutual mental models of each other's primary responsibilities. Two other items address the issues of mutual confidence and perception of mutual confidence among the crew members. The purpose of the questions is to assess the congruence of the situational and mutual mental models of crew members prior to the mission. The questionnaire includes both open-ended questions and rating scales. For the open-ended questions, we compared the responses of the two crew members and rated their similarity.

The Crew Member Post-Mission Questionnaire was completed independently by each crew member after each scenario. The items in this questionnaire examine confidence in one's partner, the extent to which each crew member felt he was able to anticipate (i.e., predict) the actions and decisions of the other, the extent to which crew members felt they acted "in sync" with each other, and the extent to which they monitored and assisted each other. The goal is to assess the congruence of situational and mutual mental models of crew members subsequent to mission performance.

Table 1 summarizes the data available on mental-model congruence. Note that the majority of the items provide indirect, not direct, evidence of the congruence of the crews' mental models. Comparison of the similarity of the crew members' responses to the open-ended questions provides a direct measure of the extent to which they had similar models of the situation and of each other. Other items dealing with mutual confidence, the perceived ability to anticipate each other's actions and decisions, and the perceived ability to act "in sync," are indirect indicators that the crew had congruent models of the situation and of each other. Questions dealing with cross-monitoring behavior and providing assistance to one's partner are even more indirect measures of mental-model congruence. We expected that crews with more-congruent mutual mental models would be better able to provide assistance to each other, but might be less likely to cross-monitor their partners.

Communication-Based Measures

We developed a recording instrument and procedures to code and analyze cockpit communications in the Battle-Rostering Experiment based on videotapes from the flight simulator. The instrument and procedures were based on the communication-analysis data-collection instrument and methodology used to analyze cockpit communications during Year 1 of this project (Entin, Entin,

Table 1

Data Available on Congruence of Crew Mental Models

Question	Purpose of question and transformation for analysis
Pre-Mission	
1. List up to 3 "show stoppers" that could compromise mission (open ended)	Assess congruence of mental models of the situation Compare responses of two crew members and code similarity on a scale of 1 to 7
2. Briefly describe your two most important tasks and responsibilities (open ended)	Assess congruence of mutual mental models Compare responses to responses of other crew member and code similarity on a scale of 1 to 5
3. Briefly describe your fellow crew member's two most important tasks and responsibilities (open ended)	Assess congruence of mutual mental models Compare responses to responses of other crew member and code similarity on a scale of 1 to 5
4. How much confidence do you place in the ability of your fellow crew member? (Rate from 1 to 7)	Assess congruence of mutual mental models
5. How much confidence do you think your fellow crew member places in you? (Rate from 1 to 7)	Assess congruence of mutual mental models
Post Mission	
1. How much confidence did you have in your fellow crew member? (Rate from 1 to 7)	Assess congruence of mutual mental models
2. How much assistance did you provide to your fellow crew member? (Rate 1 to 7)	Assess congruence of mutual mental models
3. How much did you cross-monitor the actions of your fellow crew member? (Rate 1 to 7)	Assess congruence of mutual mental models
4. To what extent were you able to anticipate the actions of your fellow crew member? (Rate 1 to 7)	Assess congruence of mutual mental models
5. To what extent were you acting in sync with your fellow crew member?(Rate 1 to 7)	Assess congruence of mutual mental models

MacMillan, & Serfaty, 1993). The methodology allows us to categorize cockpit communications at an intermediate level of granularity. Guided by our theoretical framework, our categorization system allows us to obtain important information about cockpit communication and coordination patterns without requiring us to carry out a complete content analysis of the communications. This subsection describes the data-collection instrument we are using, the procedure we are following, and the progress made to date on the communication analysis of the videotapes from the Battle-Rostering Experiment.

The Data Collection Instrument

Figure 4 shows the data collection instrument being used. The instrument is designed to capture the type, function, source, and directionality of communications between the two cockpit crew members and between them and other crew and ground personnel. The type and functionality of communications are captured in the rows of the recording instrument, and the source and directionality in the columns.

We categorize communications into three distinct types: requests, responses, and transfers. Requests include information or actions solicited by a member of the crew. Responses are communications made in answer to requests. Transfers are unsolicited statements of information or actions. An "other" category is included in the instrument in case there is a need to capture any utterances that do not fit into one of the three specified types.

Within each type of communication there are three functional areas: information, action/task, and problem solving/planning. Information utterances can be statements or requests for information. Action/task utterances can be statements of actions taken or about to be taken or requests to another person to take an action or carry out a task. The planning and problem-solving function is used for utterances that refer to future plans or relate to problem-solving. Utterances in each function category can be classified by type as a request, response, or an unsolicited transfer. Examples of utterances classified by type and function are given in Table 2.

The columns of the data collection instrument are used to capture the source and directionality of communications. The source of a communication can be the pilot, the copilot/gunner (CPG), or "other." The other category includes communications to and from other flightcrews and from the ground. The direction of the communication can be from the pilot to the CPG or to other, from the CPG to the pilot or to other, or from other to the crew. We separate communications made *to* other by the pilot from those made by the CPG, but because communications *from* other may be directed to the cockpit crew as a unit there is only one category for recording communications from other to the cockpit crew.

DRAFT

Team ___ Scenario ___ Mission ___ Time ____ to ____ Coder ___

TYPE		PILOT TO:		CPG TO:		OTHER TO:
		CPG	Other	Pilot	Other	Crew
REQUESTS	Information					
	Action/ Task					
	Problem Solving/ Planning					
RESPON- SES	Information					
	Action/ Task					
	Problem Solving/ Planning					
TRANSFERS	Information					
	Action/ Task					
	Problem Solving/ Planning					
OTHER (Specify)						

Figure 4. Data recording instrument used to code cockpit communications.

Table 2

Examples of Utterances by Type and Function

Type/Function	Information	Action/Task	Problem Solving/Planning
Request	How far is it? Do you see the targets?	Turn right. Tell me when the light goes out.	We should try to figure it out. We need to keep an eye on ...
Response	About 10 clicks out. I don't see them.	I'm turning right. The light went out.	I'm going to figure it now.
Transfer	You're clear on the left. We have a target destroyed.	I'll be coming up. I'll pull the circuit breaker.	When we get to the FARP we'll see if we can get an update. We'll check the flight instruments once we get up. We may have to reposition because ...

The communications data-collection instrument for the Battle-Rostering Experiment expands the instrument used to analyze videotapes in Year 1 of this project. One modification, the addition of the "response" type, allows us to make a distinction between unsolicited transfers and those made in response to a request. With the addition of this new type, the same utterance can be coded into different categories depending upon the context in which it is stated. For example, the statement "You are clear on the left" would be coded as a transfer of information if it is an unsolicited utterance, whereas it would be coded as a response if it followed a question such as "Am I clear on the left?"

A second modification made to the data-collection instrument provides a way to capture acknowledgments of communications as a distinct category. The shaded area of each cell in the data collection instrument is used to record acknowledgments of a

communication. An acknowledgment imparts no new information, but is a verbal acknowledgment that an utterance was heard. Most often an acknowledgment is either an "OK" or "roger" made by the recipient of the communication to the sender of the communication. Thus tally marks in the shaded area of a cell actually represent communications in the opposite direction to those made in the white (unshaded) area. For example, in the row recording transfers of information, the white area of the cells in the first column of the instrument would record transfers from the pilot to the CPG whereas the shaded area would record the CPG's acknowledgments of these information transfers from the pilot.

Data Collection Procedure

In the analysis conducted in Year 1, we coded the cockpit communications for an entire scenario. Based on that experience, we concluded that stable and useful communication measures could be obtained by coding selected time segments in a scenario, rather than the entire episode. In the Battle-Rostering Experiment, coding of the crew communications is centered around two critical events that occur during each scenario: an equipment malfunction and arrival at the battle position (BP). These are the two events for which workload ratings were taken (see discussion below). We code 16 minutes of communication in each scenario: seven minutes centered around the equipment malfunction and nine minutes around the arrival at the battle position.

For each event the time period is divided into several one- or two-minute time segments. Figure 5 shows the relationship between the collection of workload data and the communication data for each of the two critical events. The time segments immediately surrounding an event correspond to the time periods during which the workload measurements were taken. The segments at the beginning and end of the time period respectively precede and follow the periods in which the workload measurements were taken. A separate data-collection form is used for each time segment, and the starting and the ending time of the observation period are recorded at the top of the instrument. The crew number, scenario number, mission number, and rater are also recorded at the top of the form.

Planned Analysis of the Communication Measures

Two types of measures will be computed from the communication instrument: communication volumes and communication ratios. Communication volume measures count the number of utterances in a particular category. To compare the volume of communication across time periods of differing length, we will calculate communication volume per unit time, with the basic unit being one minute. Communication ratios can be used to create measures that are independent of the volume of communications (for example, the ratio of number of pilot to CPG communications or the acknowledgment ratio, defined as the number of acknowledgments divided by the number of communications) or to calculate higher-

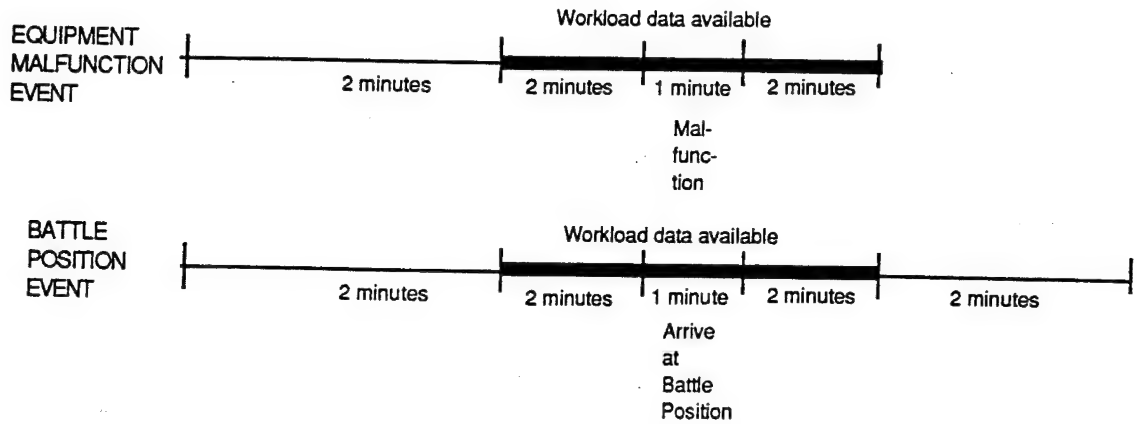


Figure 5. Relationship between time periods for communication data and workload data.

level communication measures that compare two categories, such as the anticipation ratio (ratio of information transfers to information requests). Both volume and ratio measures can be compared for a particular type and/or function of communication or they can be aggregated across categories. They can be compared within crews at different time periods (for example preceding or during a high-task-demand period) or across crews (for example battle-rostered versus mixed crews), and they can be related to other measures of crew coordination, such as those derived from questionnaires completed by instructor pilots (IPs).

Progress to Date

A preliminary coding of the crew communications was conducted by two members of the ALPHATECH staff who are proficient in protocol analysis. The videotape of one crew was randomly selected, and random time periods on the tape were used for the preliminary coding. This preliminary coding was carried out to assess the feasibility of the coding categories used on the revised data-collection instrument and to obtain a preliminary indication of inter-rater reliability.

On the basis of this preliminary coding, the two raters concluded that the expanded set of "type" categories were mutually exclusive and empirically recognizable. They also concluded that the "acknowledgment" category (in which the direction of communication is reversed) is usable and does not confuse the coder. In this preliminary coding effort the two raters were able to achieve at least 70 percent agreement in the coding of utterances. The two raters also worked together to establish standards for categorizing utterances (e.g., what distinguishes a planning utterance from an information or action utterance?).

They concluded that additional calibration work would be needed to refine the coding conventions to insure inter-rater reliability. The preliminary analysis was also used to verify that the time segments in the scenarios for which workload ratings were taken could be reliably located on the videotapes.

Both raters experienced some difficulty in distinguishing between the voices of the pilot and the CPG. Although the pilot and CPG positions are identified on the videotape, it is not possible to visually determine who is speaking because the crew members' faces are not visible. Raters must rely on voice recognition to distinguish one speaker from the other. In cases where the voices were difficult to distinguish, the raters found that the speaker could often be identified on the basis of the contents of the communication (for example, an utterance such as "I'll slow up a bit" is easily attributable to the pilot). There were some utterances, however, for which the raters were not sure of the identity of the speaker even after listening repeatedly. The raters concluded that it would be necessary to view a larger time segment of the tapes (beyond the 16 minutes to be coded) in order to train themselves to reliably distinguish one speaker from the other.

After completing this preliminary analysis of a single tape, the raters coded a subset of five tapes to obtain a reliable estimate of the amount of time required per videotape. This coding work is currently being completed. Based on results to date, it takes approximately two hours per videotape to obtain the communication measures. Approximately two-thirds of that time is required to do the coding, with the remainder of the time used in locating and identifying the time segments to be analyzed and in listening to the other portions of the tape in order to link the speakers' voices to their cockpit positions.

On the basis of the tapes that have been viewed thus far, the raters feel that they can identify the speaker with a reasonable degree of confidence for about 90 percent of the utterances. After both raters have completed the ratings for the subset of five tapes, the two raters will compare their coding work in order to see whether they are consistent in the identification of speakers. If the inter-rater reliability is acceptable, we will maintain the distinction between speakers; if it does not reach an acceptable level, it will be necessary to drop this distinction.

Once the preliminary analysis of inter-rater reliability is complete, we will code the entire set of 48 tapes. In order to calculate a final inter-rater reliability measure, both raters will analyze a common set of eight (approximately 20%) of the tapes. With this overlap, we estimate that the coding task will take approximately 112 person-hours to complete. Once all of the tapes have been coded, we will compute and analyze the communication measures and integrate the results with the data analysis already conducted.

Ratings of Teamwork

The Instructor Pilot (IP) Post-Mission Teamwork Rating Form, was designed to measure detailed teamwork behaviors expected to be sensitive to the congruence of the crew's mental models. The final form of the IP Post-Mission Teamwork Rating Form is included in Appendix B. The questionnaire was completed after each mission by the IPs, who were domain experts. The 16 items of the instrument are constructed around six dimensions of teamwork and various aspects of crew members' mental models, including the extent to which each crew member anticipates the other's actions and decisions, the extent to which each crew member understands the actions and decisions of the other, the extent to which crew members share workload, and the extent to which crew members support one another. Each item in the questionnaire is accompanied by a seven-point scale, with behavioral anchors that describe the two ends of the scale.

In addition to ratings from the IP Post-Mission Teamwork Rating Form, we obtained from DRC the ratings from the Aircrew Evaluation Checklist (ACE) also completed by the IPs. The ACE was developed by Simon (1991) to support the training of Army aviation crews by providing an assessment of the quality of the crew's coordination. The ACE "measures an aircrew's ability to integrate a variety of human factors principles into the cockpit milieu" (Simon, 1991, p. 95). We use the average ACE score as an overall indicator of the quality of the crew's coordination and teamwork.

Workload Measures

The measurement of crew workload during mission performance in the Battle-Rostering Experiment was challenging because it was not feasible to interrupt crews in mid-flight in order to take workload measures. Instead, workload ratings were obtained using an unusual retrospective method in which videotapes were used to facilitate recall. After each scenario, crew members viewed short videotape segments of themselves during the scenario and rated their workload, as they remembered it, during each segment. Crew members were asked to complete the Task Load Index (TLX) workload-assessment questionnaire for each segment. The TLX is a well-documented subjective workload assessment instrument developed at NASA (Hart & Staveland, 1988). The TLX calls for six ratings of different aspects of workload (mental demand, physical demand, time pressure, performance, effort, and frustration), each rated on a scale of 1 to 20.

Workload was rated for six time segments in each scenario, tied to events that occurred in all four scenarios:

Event 1: Equipment Malfunction (5 minutes)

Time 1: Two minutes prior to malfunction
Time 2: During malfunction (approximately one minute)
Time 3: Two minutes after malfunction

Event 2: Battle Position (5 minutes)

Time 1: Two minutes prior to arriving at battle position
Time 2: Arrival at battle position (approximately one minute)
Time 3: Two minutes after arriving at battle position

In the equipment malfunction event, we expected that the situation would be most demanding at Time 2 and would begin to be less demanding during Time 3. In the battle position event, in contrast, we expected that the situation would become more demanding at Time 2, but might become most demanding at Time 3. Overall, we expected that the battle position event would generate higher perceived workload than the equipment malfunction because it involves engaging the enemy and occurs at a location where it is possible for the helicopter to come under enemy fire.

Results show that perceived workload was, in fact, significantly higher during the battle position event than during the equipment malfunction event (The mean TLX scores were 8.74 and 7.70, respectively, $F=6.48$, $p=.020$). This indicates that the retrospective workload ratings were sensitive to differences in the scenario events. Furthermore, there is a significant interaction between event and time period within the event, i.e., the three measurement time periods did not follow the same pattern for each event. As explained above, we expected that workload would decrease between Time 2 and Time 3 in the equipment malfunction event, but would be stable or even increase between Times 2 and 3 in the battle position event. In fact, this is the pattern found in the mean workload ratings, as shown in Figure 6. We conclude that our workload rating technique based on post-scenario videotape observation produced results that were sensitive to events in the scenario, and that the relationship between scenario events and changes in perceived workload seems reasonable.

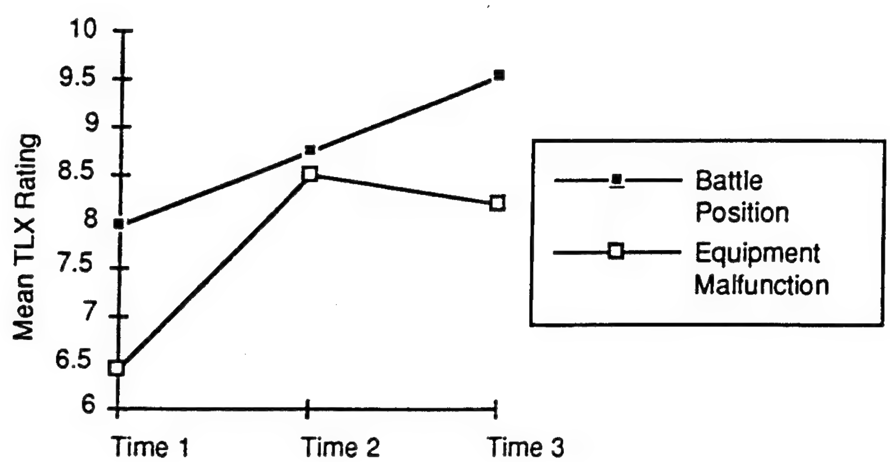


Figure 6. Sensitivity of TLX ratings to scenario events.

Ratings of Crew Performance

The primary source of crew performance data for the Battle-Rostering Experiment are ratings obtained from DRC, reported more fully in Grubb, Simon, Leedom, and Zeller (1994). Two types of evaluation measures that assess crew effectiveness were provided by DRC. The first measure was an overall grade, assigned by the IP and based on the IP's overall assessment of the crew's performance on the mission. The grade can have one of four values: U, S-, S, and S+ (coded as 0, 1, 2, and 3, respectively), where U stands for Unsatisfactory and S for Satisfactory. The second performance-evaluation measure was a set of gradeslips based on 25 Aircrew Training Manual (ATM) measures. The ATM measures assess a crew's performance on specified tasks. The gradeslips were filled out by the IPs, who used the same evaluation scale used for the overall grade. We use an average of the ATM measures for analysis.

In order to explore the relationship between workload and performance, we also asked the IPs to rate crew performance in each scenario segment for which workload data were collected. These ratings used the same scale used for the overall grade.[2] These post-mission ratings for each crew provide more-detailed data on performance during the periods in which workload was also measured. These performance-rating questionnaires may be found in Appendix C.

Summary of Data Available for Analysis

Figure 7 summarizes the data available from the Battle-Rostering Experiment to test the hypothesized relationships between crew mental-model congruence and crew performance that were shown in Figure 3. Mental-model congruence measures come from the pre- and post-mission questionnaires administered to the crew. Crew-communication and coordination measures come from videotape analysis.[3] Ratings of teamwork quality come primarily from the post-mission questionnaire completed by the IPs, supplemented by the crew's total score on the ACE. Workload measures are obtained from the TLX ratings made by crew members, for six different portions of the mission. The primary crew-performance measures (supplied by the IPs) are the crew's overall grade and the average of the ATM-based measures, supplemented by the more-detailed segment-by-segment performance ratings that correspond to the time segments for which workload data are available.

---

[2] One of the performance measurement points ("during the battle position") corresponds to two of the workload measurement points (entering the battle position and two minutes afterwards).

[3] Analysis of the Battle-Rostering Experiment videotapes is not yet complete, and the results are not included in this interim report.

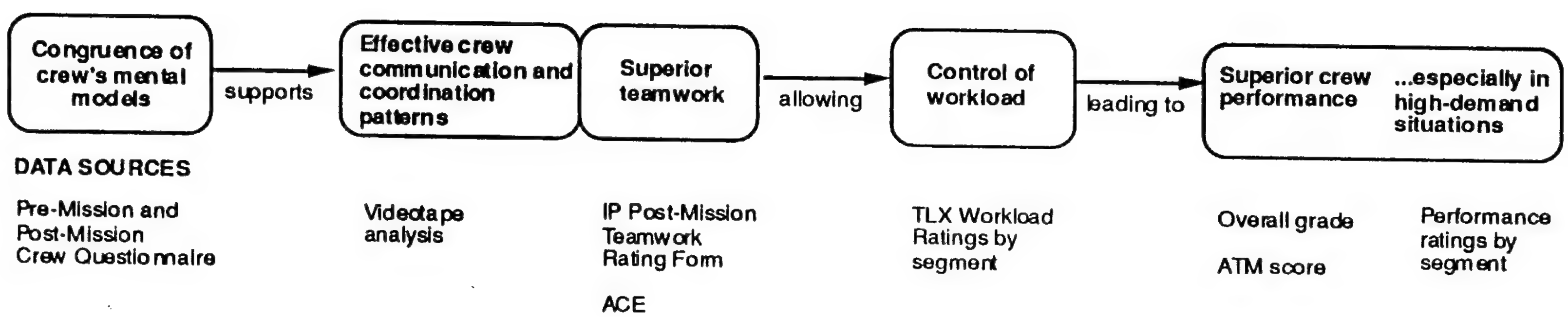


Figure 7. Data available to test hypothesized relationships.

Results: Mental-Model Congruence

Our analysis of the results of the Battle-Rostering Experiment systematically tests the theoretical links shown in Figure 3. The first question for analysis is whether battle rostering or crew-coordination training was associated with more congruent mental models among crew members. Two aspects of mental model congruence were examined: the congruence of the crew's mental models of the situation, and the congruence of their mutual mental models of each other.

Table 3 presents responses to the Pre- and Post-Mission Questionnaires for battle-rostered crews and mixed crews. The table shows a number of significant differences in the perceptions of individuals serving in battle-rostered crews and mixed crews. For items that combine the responses of the two crew members (the first three items in the Pre-Mission Questionnaire), the analysis is a t-test comparing mixed versus battle-rostered crews. For the individual responses (the remaining items), the analysis is a paired-comparison t-test of differences in the perceptions of the same individual while serving in a battle-rostered crew and in a mixed crew.

Before the mission was flown, crew members had more confidence in their battle-rostered partner's ability than in their mixed-crew partner's ability. They also felt that their partner had more pre-mission confidence in them in the battle-rostered than in the mixed-crew condition. The measures based on our comparison of the crew members' responses to the open-ended questions showed no significant differences between the battle-rostered and mixed crews, however. The battle-rostered crews felt more confidence in their partners, but there is no evidence that they had more congruent shared perceptions about the situation or each other's roles before the mission.

Crew members continued to have more confidence in their partner's abilities in the battle-rostered condition after the mission had been flown. They felt that they were better able to anticipate the actions and decisions of their battle-rostered partner, consistent with the hypothesis that battle-rostering promotes the development of mutual mental models that allow crew members to anticipate each other's needs. Surprisingly, the post-mission level of confidence in one's partner was somewhat lower than the pre-mission level in both the battle-rostered and the mixed conditions.

Crew members felt they provided *less* assistance to their partner in the battle-rostered than in the mixed condition. There also seems to be a tendency for crew members to do less cross-monitoring in the battle-rostered condition, although the difference is statistically significant at the $p<.10$ level only if a one-tailed test is used. Both of these results support the hypothesis that battle rostering may induce some complacency into

Table 3

Mental-Model Congruence for Battle-Rostered and Mixed Crews (Pre-Mission and Post-Mission Questionnaires)

Questionnaire Items	**BR Condition** Mean	**Mixed Condition** Mean	t value
Pre-Mission Questionnaire			
Congruence of 3 show stoppers that could compromise mission (1 to 7)	3.81	4.02	-.37 (n=12)
Congruence of pilot's two most important tasks and responsibilities (1 to 5)	2.85	2.89	-.14 (n=12)
Congruence of copilot's two most important tasks and responsibilities (1 to 5)	3.17	3.29	-.41 (n=12)
How much confidence do you place in the ability of your fellow crew member? (1 to 7)	6.43	6.19	2.08* (n=24)
How much confidence do you think your fellow crew member places in your ability? (1 to 7)	5.94	5.50	3.23** (n=24)
Post-Mission Questionnaire			
How much confidence did you have in your fellow crew member? (1 to 7)	5.96	5.59	1.95‡ (n=24)
To what extent were you able to anticipate the actions and decisions of your fellow crew member? (1 to 7)	5.67	4.98	3.34** (n=24)
How much assistance did you provide your fellow crew member? (1 to 7)	3.88	4.28	-1.95‡ (n=24)
How much did you cross monitor the actions of your fellow crew member? (1 to 7)	3.79	4.12	-1.44 (n=24)
To what extent were you acting "in sync" with your fellow crew member? (1 to 7)	5.02	4.77	.88

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

the crew's teamwork behavior. Perhaps the battle-rostered crews felt less need to assist their partner because they had higher confidence in him, and felt that they could anticipate the needs of their partner based on their previous experience without performing frequent cross-monitoring.

In the pilot experiment conducted in June 1993, initial versions of the crew questionnaires were tested. In this experiment the pre- and post-mission questionnaires were administered before and after crew-coordination training was given to the crews. Note that all crews in this pilot experiment were battle rostered.[4] Table 4 present the results of the Post-Mission Questionnaire administered before and after training. An examination of the results provides an indication of whether the crew-coordination training improved mental-model congruence over and above the effects of battle rostering.

Post-mission questionnaire responses indicate that the coordination training provided to crews increased the congruence of the crew's mutual mental models. Crew members reported, post mission, that they had more confidence in their partner after training, and felt that they were better able to anticipate his actions and decisions. Coordination training also increased the extent to which the crew members perceived that they performed cross-monitoring of their partners. This is in contrast to the effect of battle rostering on perceived cross-monitoring behavior, which was negative. The effect of training on the amount of assistance that crew members felt they gave to their partners was not significant.

Comparing the effects of battle rostering and crew-coordination training (in addition to battle rostering) on the congruence of the crews' mutual mental models indicates that both battle rostering and training increase the crew members' feelings of confidence in each other and their sense that they can anticipate their partner's actions and decisions, i.e., increase the crew members' perception that they have congruent mutual mental models. Training appears to increase perceived cross-monitoring behavior, while battle rostering is associated with lower levels of perceived cross-monitoring. Also, coordination training had no effect on the amount of assistance that crew members felt they provided to their partners, while battle rostering was associated with the perception that less assistance was provided.

Results: Teamwork Ratings

The next step in the analysis is to examine the IP ratings of the quality of crews' teamwork, and determine whether more-

[4] There were 14 crews in the June pilot experiment, but only eight of these were "real" crews, i.e., crews that did not include an IP. The results presented here are based only on these eight battle-rostered crews.

Table 4

Mental-Model Congruence Before and After Crew-Coordination Training (Post-Mission Questionnaires)

Questionnaire Items	**Before Coordination Training** Mean	**After Coordination Training** Mean	t value (n=16)
Post-Mission Questionnaire			
How much confidence did you have in your fellow crew member? (1 to 7)	5.25	5.81	-2.76*
To what extent were you able to anticipate the actions and decisions of your fellow crew member? (1 to 7)	5.13	5.69	-2.52*
How much assistance did you provide your fellow crew member? (1 to 7)	4.25	4.37	-.56
How much did you cross-monitor the actions of your fellow crew member? (1 to 7)	3.75	4.31	-1.95‡
To what extent were you acting "in sync" with your fellow crew member? (1 to 7)	4.91	5.69	-1.74

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

congruent mental models were associated with better teamwork as observed by the IPs, and whether battle rostering or crew-coordination training improved the crew's teamwork ratings.

The IP Post-Mission Teamwork Rating Form asked IPs to rate 16 aspects of the crews' coordination behavior (see Appendix B). Not one of these 16 ratings was significantly different for the battle-rostered versus mixed crews, however. The IP ratings provide no evidence of superior teamwork on the part of the battle-rostered crews. There are two possible conclusions—either the instrument was not sensitive enough to detect teamwork differences in the battle-rostered and mixed crews, or the crews' coordination performance (from the point of view of the IPs) was the same whether they were battle rostered or mixed.

The pretest of the IP Post-Mission Teamwork Rating Form conducted as part of the June pilot experiment shows many significant differences in teamwork before and after the crews were trained, as shown in Table 5. After crew-coordination training, the crews were rated as communicating better, showing more anticipation of each other's needs, alerting each other more frequently, monitoring each other's behavior, and providing feedback and backup. The sensitivity of the questionnaire to pre- and post-training differences indicates that the failure to find differences between the battle-rostered and mixed crews may be because there were no substantial differences to find, not because the questionnaire lacked the sensitivity to detect differences.

Relationship of Teamwork to Mental-Model Congruence

One of our hypotheses is that more-congruent mental models support the communication patterns associated with superior teamwork. We can test this hypothesis by examining the correlations between the crew members' responses to the Pre- and Post-Mission Questionnaires (averaged for each crew) and the IP Post-Mission Teamwork Rating Form teamwork ratings for each crew. Table 6 presents these results for the Battle-Rostering Experiment. Ratings based on the three open-ended items from the Pre-Mission Crew Questionnaire have been omitted because they were not found to be significantly related to the teamwork ratings.

Table 6 shows that the questionnaire responses of the crew members both before and after the mission were highly correlated with the 16 ratings of teamwork provided by the IPs. Confidence in one's partner, both before and after the mission, was significantly correlated with a greater orientation toward teamwork, fewer communication and individual errors, better monitoring of each other's behavior, better provision of alerts, better feedback, better backup, better adjustment of responsibilities, better overall coordination, and a more congruent understanding of the mission. After the mission, confidence in one's partner was also significantly correlated with acknowledgments and with the anticipatory provision of information and assistance. These results may be interpreted in two ways. First, crews that had more confidence in each other before the mission was flown showed better teamwork during the mission. Second, crews that gave more acknowledgments and showed more anticipatory behavior during the mission had higher confidence in each other after the mission was completed than crews that gave fewer acknowledgments and exhibited less anticipatory behavior.

Crew responses to questions dealing with cross-monitoring and providing assistance to one's partner show that the crew members apparently interpreted these items in a negative way, perhaps as indicating that they lacked confidence in their partner. The correlation between pre-mission confidence in one's partner and the provision of assistance to one's partner during the mission (from the Post-Mission Questionnaire) was -.54 ($p<.01$) and the correlation of confidence with reported cross-monitoring was -.58

Table 5

IP Teamwork Ratings Before and After Coordination Training

IP Teamwork Ratings (1 to 7)	**Before Training** Mean	**After Training** Mean	t value (n=8)
Crew oriented toward teamwork?	4.50	4.37	.36
Errors caused by inadequate communication?	3.50	4.00	-1.08
Errors caused by inadequate individual actions?	3.50	4.50	-2.00‡
How well did crew members communicate?	3.38	4.00	-2.38*
How well did crew members acknowledge other's messages?	3.12	3.87	-1.43
CPG provide relevant information without being asked?	3.13	3.88	-2.39*
Pilot provide relevant information without being asked?	3.13	4.25	-2.55*
Crew members monitor each other's behavior?	3.13	4.25	-2.55*
Crew members alert each other to impending decisions and actions?	2.88	4.12	-5.00**
Crews provide feedback to each other?	3.50	4.13	-1.93‡
Crews provide backup to each other?	3.38	4.25	-2.97*
Pilot anticipate the need to provide assistance to the CPG?	3.50	4.13	-2.38*
CPG anticipate the need to provide assistance to the pilot?	2.75	4.00	-5.00**
Crew members adjust responsibilities to prevent overload?	3.38	4.13	-3.00*
Were the crew's behaviors coordinated?	3.38	4.25	-3.87**
Pilot's and CPG's understanding of the mission congruent?	4.50	4.87	-1.00

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

DRAFT

Table 6

Correlation of Crew Mental-Model Congruence with IP Teamwork Ratings

	Correlation with Measures of Mental-Model Congruence from Crew Questionnaires (n=48)						
	Pre-Mission		**Post-Mission**				
IP Ratings of Teamwork	Confidence in partner	Partner confidence in you	Confidence in partner	Provide assistance	Cross-monitor	Anticipate	Act in sync
Crew oriented toward teamwork?	.34*	.25*	.21‡	-.44**	-.31*	.13	.23‡
Errors caused by inadequate communication?	.34*	.23‡	.29*	-.45**	-.39**	.37**	.36**
Errors caused by inadequate individual actions?	.21‡	.25*	.24*	-.22‡	-.16	.34**	.32*
How well did crew members communicate?	.18	.17	.15	-.28*-	-.11	.08	.23‡
How well did crew members acknowledge other's messages?	.16	-.02	.23‡	-.28*	-.29*	.07	.21‡
CPG provide relevant information without being asked?	.16	-.02	.35**	-.30*	-.28*	.30*	.29*
Pilot provide relevant information without being asked?	.18	.15	.29*	-.38**	-.26*	.19‡	.11
Crew members monitor each other's behavior?	.26*	.18	.27*	-.43**	-.38**	.26*	.29*
Crew members alert each other to impending decisions and actions?	.32*	.14	.21‡	-.36**	-.28*	.42**	.26*
Crews provide feedback to each other?	.28*	.19‡	.21‡	-.34**	-.23‡	.17	.17
Crews provide backup to each other?	.36**	.18	.32**	-.37**	-.26*	.37**	.31*
Pilot anticipate the need to provide assistance to the CPG?	.18	-.01	.44**	-.32*	-.28*	.32*	.28*
CPG anticipate the need to provide assistance to the pilot?	.12	.12	.23‡	-.31*	-.17	.39**	.11
Crew members adjust responsibilities to prevent overload?	.35**	.35**	.36**	-.42**	-.36**	.33**	.28*
Were the crew's behaviors coordinated?	.31*	.31*	.21‡	-.36**	-.28*	.42**	.35**
Pilot's and CPG's understanding of the mission congruent?	.21‡	.29*	.20‡	-.29*	-.09	.23‡	.15

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

(p<.01). Provision of assistance and cross-monitoring, as measured by the Post-Mission Questionnaire, were significantly and negatively correlated with almost all of the IP teamwork ratings, including the IP's ratings of whether the team cross-monitored each other and whether they anticipated the need to provide assistance to each other! The results indicate that, while the IPs perceived cross-monitoring and the provision of assistance when needed as associated with "good" teamwork, the aviators had the opposite perception and, if they had confidence in their partner, reported that they did not monitor his behavior or assist him.

Crew members' perceptions that they could anticipate their partner's actions and decisions and were able to think and act in sync with him were positively correlated with many of dimensions of teamwork rated by the IPs. Crews that reported a better ability to anticipate and act in sync were rated as having more teamwork orientation, making fewer errors, anticipating each other's needs for information and assistance, providing more alerts and backup, adjusting responsibilities to prevent overload, and having better overall coordination.

Examination of the correlations between crew questionnaire responses and ratings made by the IPs using the ACE instrument confirm the results found using the IP Post-Mission Teamwork Rating Form. Table 7 shows that the average ACE score for teams was significantly and positively correlated with crew members' confidence in one another, their perception that they were able to anticipate each other's actions and decisions, and their perception that they were in sync with their partner. The average ACE score was negatively correlated with reports of cross-monitoring and providing assistance to one's partner.

Table 7

Correlation of Measures of Crew Mental-Model Congruence with Average ACE Rating

	Correlation with Measures of Mental-Model Congruence from Crew Questionnaires (n=48)						
	Pre-Mission		**Post-Mission**				
	Confidence in partner	Partner confidence in you	Confidence in partner	Provide assistance	Cross-monitor	Anticipate	Act in sync
Average ACE Rating	.30*	.09	.34**	-.43**	-.32*	.30*	.34**

‡ p ≤ .10
* p ≤ .05
** p ≤ .01

Results: Workload Ratings

Our theory suggests that more-effective coordination and better teamwork allow a crew to keep their workload within manageable levels even when task demands are high, resulting in better crew performance. In order to test this hypothesis, we collected subjective workload ratings using the TLX instrument at six points in each scenario, as described above.

Workload was measured during the Battle-Rostering Experiment, but not during the pilot experiment, so that we have results that enable us to compare battle-rostered with mixed crews, but no results that allow us to assess the effect of coordination training on workload. The comparison of battle-rostered and mixed crews shows that perceived workload was significantly higher in the battle-rostered condition. The mean TLX score was 8.41 for battle-rostered crews and 8.03 for mixed crews ($F= 4.83$, $p=.041$). Apparently the crew members felt that they were working harder when they flew with their battle-rostered partner.

Teamwork and Workload

Our theory predicts that better teamwork, as measured by the IP Post-Mission Teamwork Rating Form and the ACE, should be negatively correlated with perceived workload as rated on the TLX. It also suggests that lower perceived workload should be positively correlated with crew performance. This positive correlation should be especially strong during the high task-demand portions of the scenario, e.g., during equipment malfunctions and while the crew was in battle position.

Table 8 shows the correlation between crew workload, averaged across the six measurement points in each scenario, and the 16 teamwork ratings from the IP Post-Mission Teamwork Rating Form as well as the correlation of crew workload with the average ACE score. The results show that a number of teamwork dimensions were associated with lower perceived workload on the part of the crew. Lower perceived workload was significantly correlated with fewer communication errors, better communication, better acknowledgments, more monitoring of each others' behavior, more alerts, more feedback, more anticipation of partner's needs for assistance, and more adjustment of responsibilities to prevent overload. As predicted by the theory, the crews that were rated as maintaining better communications, monitoring each other's behavior, anticipating each other's needs, and adjusting responsibilities as needed were able to keep their workload at a lower level. The average ACE score was also significantly negatively correlated with perceived workload, confirming that, overall, better teamwork is associated with a lower perceived workload.

Table 8

Correlation of Average Crew Workload with IP Teamwork Ratings and Average ACE Rating

	Correlation with Team Workload (Average TLX Score) (n=48)
IP Ratings of Teamwork	
Crew oriented toward teamwork?	-.05
Errors caused by inadequate communication?	-.19‡
Errors caused by inadequate individual actions?	-.14
How well did crew members communicate?	-.24*
How well did crew members acknowledge other's messages?	-.25*
CPG provide relevant information without being asked?	-.14
Pilot provide relevant information without being asked?	-.18
Crew members monitor each other's behavior?	-.33**
Crew members alert each other to impending decisions and actions?	-.23‡
Crews provide feedback to each other?	-.44**
Crews provide backup to each other?	-.16
Pilot anticipate the need to provide assistance to the CPG?	-.18
CPG anticipate the need to provide assistance to the pilot?	-.37**
Crew members adjust responsibilities to prevent overload?	-.27*
Were the crew's behaviors coordinated?	-.14
Pilot's and CPG's understanding of the mission congruent?	-.03
Average ACE Rating	-.25*

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

Workload and Performance

Our theory also predicts that crews with lower perceived workload will achieve better performance, especially in high-demand situations. Table 9 presents the correlations between perceived workload (averaged across the six measurement points in each scenario) and overall crew performance as measured by the overall grade and the average ATM score provided by the IPs. The table shows that the ATM score for the crews, which averages the grades given by the IPs on 25 different tasks specified in the Aircrew Training Manual, is significantly negatively correlated with average perceived workload—as predicted by the theory. The correlation of workload and the overall grade for the crew, while negative, is not significant.

Table 9

Correlation of Average Crew Workload with Overall Crew Performance

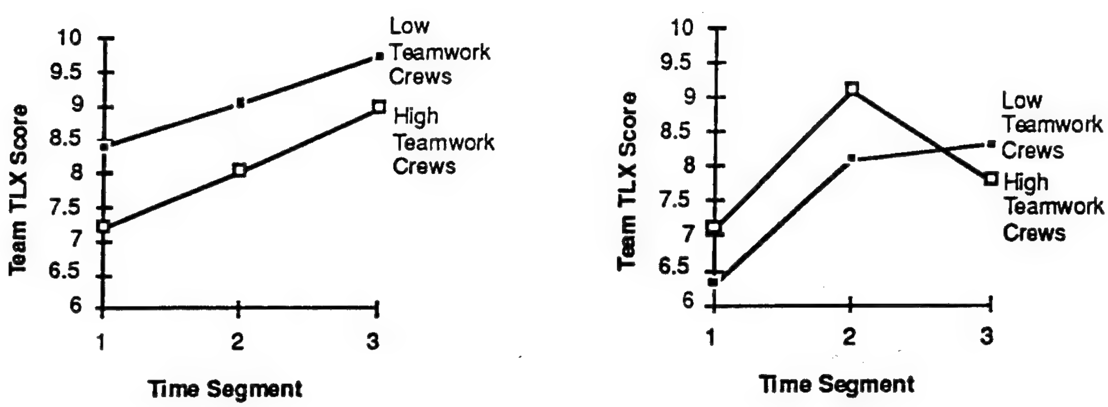
	Correlation with Team Workload (Average TLX Score) (n=48)
Crew Performance Measures	
Overall Grade	-.13
Average ATM Score	-.27*

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

Our theory predicts that the relationship between workload and performance will be especially important in high-demand situations, when the crew's ability to keep its workload within manageable limits is especially critical to its performance. We expected that teams exhibiting better teamwork would be able to control their workload more effectively in high-demand situations and, therefore, would be able to achieve a better level of performance. In order to obtain data to test this hypothesis, we asked the IPs to rate the performance of each crew specifically for those segments of the scenario for which workload ratings were also obtained. For the six segments rated, three were considered to be high-demand situations: the segment in which the equipment malfunction occurred, the segment in which the crew arrived at the battle position, and the segment that occurred immediately after the crew arrived at the battle position (see Figure 6).

We expected to find a relationship between the quality of the crew's teamwork, their perceived workload, and their performance in the high-demand segments of the scenario. The first step in testing this hypothesis was to identify the crews with the highest and lowest teamwork skills, based on the ratings provided by the IPs on the Post-Mission Questionnaire. There were 48 "trials" in the Battle-Rostering Experiment, where a trial is defined as a two-person crew flying one scenario. When we examined the distribution of average scores on the 16 items of the IP Post-Mission Teamwork Rating Form, we were able to clearly identify 18 trials in which crews were at the lower end of the distribution (an average score of less than 4.5 out of a maximum of 7) and 18 trials in which crews were at the upper end of the distribution (an average score of 5 or greater out of a maximum of 7). The 12 trials in the middle of the distribution were excluded to obtain a clearer contrast between high-teamwork and low-teamwork crews.

When we examine subjective workload ratings and their associated performance ratings for the six time segments, we see clear differences between the low-teamwork and the high-teamwork crews. Figure 8 shows the team workload for the malfunction and battle position events at three time points for the low-teamwork and high teamwork crews, and Figure 9 shows the equivalent results for the segment-by-segment IP performance ratings.[5]

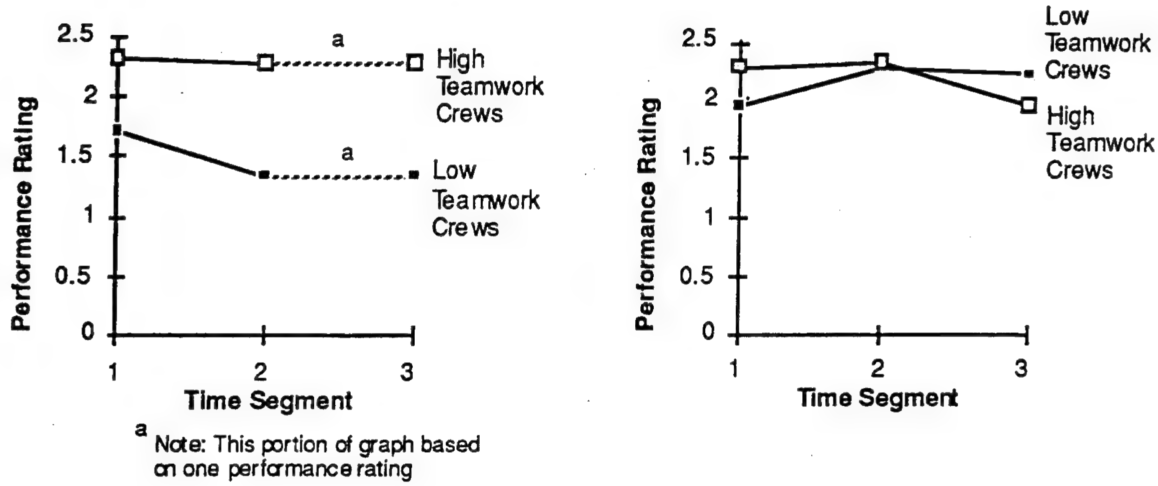


(a) Battle Position Event (n=18 per group)

(b) Malfunction Event (n=18 per group)

Figure 8. Team workload scores by scenario segment for battle position and malfunction events.

[5]The IPs were asked to provide a performance ratings for crews "during the battle position operation." This single performance rating corresponds to two workload ratings for the battle position event: arriving at the battle position (time 2) and two minutes afterwards (time 3).



(a) Battle Position Event
(n=18 per group)

(b) Malfunction Event
(n=18 per group)

Figure 9. Team performance scores by scenario segment for battle position and malfunction events.

The low-teamwork crews reported a higher subjective workload than the high-teamwork crews as they approached and arrived at the battle position.[6] The perceived workload for both groups increased over time as they arrived at the battle position.[7] The pattern of results for the malfunction event is quite different from the battle position. For low-teamwork crews, workload during the malfunction increased when the malfunction occurred (time 2) and did not decrease immediately afterwards (time 3). The high-teamwork crews, in contrast, show more variability in their workload during the malfunction. Their workload peaked sharply during the malfunction (time 2) and then immediately decreased. Overall, there was no significant difference between the two groups in their reported workload in the period surrounding the malfunction.

The performance results shown in Figure 9 are also quite different for the battle position and malfunction events. The low-teamwork crews performed less well than the high-teamwork

[6]Planned comparisons of the mean TLX scores for the two groups in each time segment show that the difference between the groups is significant at time 1 ($p<.05$), as they approached the battle position, and at time 2 ($p<.10$), as they arrived at the BP. There was no significant difference at time 3, two minutes later.

[7]Differences among the three time periods were significant, $p<.01$.

crews before and during battle position operations.[8] There was no difference in the performance of the two groups during the malfunction. When we compare the two events, we see that the high-teamwork crews performed at about the same levels at the battle position as they did during the malfunction, while the low-teamwork crews performed less well at the battle position than during the malfunction.

Comparing the results shown in Figures 8 and 9, we conclude that the high-teamwork crews showed better performance with a lower workload at the battle position. The low- and high-teamwork crews performed equally well in the period surrounding the malfunction, with no difference in their overall workload but more variability in workload for the high-teamwork crews. We did not find, as we expected, that the high-teamwork teams kept their workload more constant than the low-teamwork teams as task demand increased. Rather, the pattern of results can be more accurately described as showing that the high-teamwork crews allocated their effort more *effectively*, working hard when it was most needed (at the battle position and when the malfunction occurred), and not so hard during less critical periods (approaching the battle position and after the malfunction had been handled). They seem to have allocated their effort more intelligently, allowing them to maintain a better (more-constant) overall level of performance. The low-teamwork crews seem to have been working harder, but to less effect, as they approached and arrived at the battle position.

Results: Mental Models and Performance

The analysis reported above has linked the congruence of the crew's mental models to their performance indirectly, looking at the relationships between model congruence and teamwork, teamwork and workload, and workload and performance. This section examines the direct link between mental-model congruence, as measured by the crew questionnaires, and crew performance, as measured by the crew's overall grade and the average ATM score. Table 10 shows the correlations between these measures.

Table 10 shows no significant correlations between the crew questionnaire responses and the overall grade for the crew, but a number of significant correlations with the average ATM score. Higher confidence in one's partner both before and after the mission was associated with higher ATM scores, as was the perceived ability to anticipate one's partner's actions and decisions and the sense of acting in sync with one's partner. Reports of cross-monitoring and providing assistance to one's partner were negatively correlated with ATM score, showing the same pattern found for measures of teamwork. The higher-

[8]Differences in the performance of the two groups were significant ($p<.01$) at both measurement points.

Table 10

Correlation of Mental-Model Congruence (Pre-Mission and Post-Mission Questionnaires) with Crew Performance

Questionnaire Items	**Correlation with Overall Grade** (n=48)	**Correlation with Average ATM Score** (n=48)
Pre-Mission Questionnaire		
How much confidence do you place in the ability of your fellow crew member? (1 to 7)	.18	.21‡
How much confidence do you think your fellow crew member places in your ability? (1 to 7)	.16	.17
Post-Mission Questionnaire		
How much confidence did you have in your fellow crew member? (1 to 7)	.14	.23‡
To what extent were you able to anticipate the actions and decisions of your fellow crew member? (1 to 7)	.15	.24*
How much assistance did you provide your fellow crew member? (1 to 7)	-.14	-.39**
How much did you cross monitor the actions of your fellow crew member? (1 to 7)	-.10	-.25*
To what extent were you acting "in sync" with your fellow crew member? (1 to 7)	.18	.25*

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

performing crews were less likely to perceive (or at least to report) that they monitored or assisted each other.

Results: Teamwork and Performance

The analysis above examined the relationship between teamwork and workload, and between workload and performance. Table 11 presents the direct correlations between teamwork, as measured by the 16 items in the IP Post-Mission Teamwork Rating Form, and performance, as measured by the crew's overall grade and average ATM score. The table shows extremely high correlations between all of the teamwork measures from the IP Post-Mission Teamwork Rating Form and crew performance. The only teamwork measure that failed to be significantly correlated with both of the performance measures was the IP's rating of the congruence of the pilot's and the CPG's understanding of the mission, which was correlated only with the ATM score. We conclude that all of the dimensions of teamwork rated on the IP Post-Mission Teamwork Rating Form made a significant contribution to the crew's ability to perform their mission.

Summary of Relationships Found

Figure 10 summarizes the relationships that have been found in analyzing the results of the Battle-Rostering Experiment and its associated pilot experiment. The figure builds on the theoretical framework of Figure 3, highlighting arrows where a significant relationship was found and removing arrows where no significant relationship was found. Note that the communication analysis based on videotapes has not yet been completed, so these results are not included in the figure. Also, two relationships found in previous studies (Entin, Entin, MacMillan, & Serfaty, 1993; Simon & Grubb, 1993) have been added to the figure for completeness.

Based on responses to the Pre- and Post-Mission Crew Member Questionnaires, we found that the congruence of the crew's mental models was significantly related to their level of teamwork and to their performance. We found effects of both battle rostering and crew-coordination training on mental-model congruence. Note that our measures of mental-model congruence are indirect, however, and are based on the perception of the crew members that they could anticipate their partner's actions and decisions and could think and act in sync with their partner. Crew members' confidence in their partners was clearly related to teamwork and performance. While a high level of confidence between partners is consistent with congruent mutual mental models, it does not necessarily establish the existence of such models. The more-direct measures of mental-model congruence that rated the level of agreement between crew members in their responses to open-ended questions did not yield any significant results in the analysis.

Table 11

Correlation of IP Teamwork Ratings with Crew Performance

IP Ratings of Teamwork	Correlation with Overall Grade (n=48)	Correlation with Average ATM Score (n=48)
Crew oriented toward teamwork?	.28*	.61**
Errors caused by inadequate communication?	.45**	.73**
Errors caused by inadequate individual actions?	.71**	.45**
How well did crew members communicate?	.43**	.72**
How well did crew members acknowledge other's messages?	.36**	.62**
CPG provide relevant information without being asked?	.21‡	.44**
Pilot provide relevant information without being asked?	.37**	.74**
Crew members monitor each other's behavior?	.45**	.75**
Crew members alert each other to impending decisions and actions?	.38**	.64**
Crews provide feedback to each other?	.21‡	.60**
Crews provide backup to each other?	.27*	.57**
Pilot anticipate the need to provide assistance to the CPG?	.31*	.51**
CPG anticipate the need to provide assistance to the pilot?	.23*	.60**
Crew members adjust responsibilities to prevent overload?	.34**	.58**
Were the crew's behaviors coordinated?	.43**	.67**
Pilot's and CPG's understanding of the mission congruent?	.11	.47**

‡ $p \leq .10$
* $p \leq .05$
** $p \leq .01$

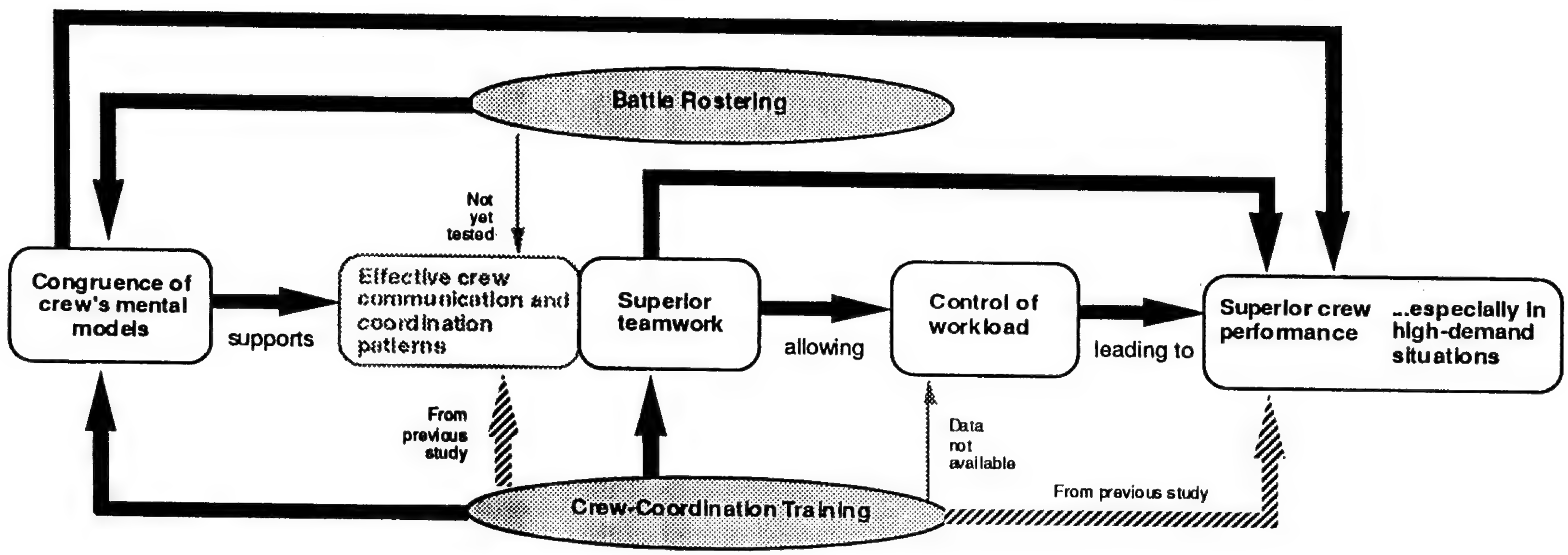


Figure 10. Relationships found between congruence of mental models, teamwork, workload, and performance.

We conclude, at a minimum, that crew members were accurately aware of whether they were able to anticipate their partner's needs and of the extent to which they were acting in sync with their partner. The perceptions of the crew as elicited by the questionnaires and the ratings made by the IPs for similar items are highly correlated.

Our assessment of the quality of the coordination and teamwork of each crew was based on a series of items completed by the IPs after the mission was completed. The measures on this IP Post-Mission Teamwork Rating Form proved to be positively correlated with mental-model congruence and negatively correlated with crew workload, supporting the hypothesis that congruent mental models provide a mechanism for superior coordination (better teamwork), allowing the crew to keep its workload at a manageable level. Further analysis of these relationships will be conducted when the videotape-based communications measures become available.

Lower perceived workload was associated with better crew performance, again supporting our theory. We also found a strong direct correlation between our teamwork measures and the overall performance of the crew as rated by the IPs. More fine-grained time-segment analysis of the relationship between teamwork, perceived workload, and performance showed that the higher-teamwork crews appear to work more efficiently, reserving their higher workload levels for high-task-demand situations. This strategy apparently allows the high-teamwork crews to maintain their performance at a reasonable level even in high-demand situations such as battle position operations. Crews with lower teamwork ratings seem to expend their effort less efficiently. Their workload ratings are higher before they reach the battle position, but they show lower performance during battle position operations.

The analysis showed little effect from battle rostering, except to increase the crew members' confidence in each other. Higher confidence was associated with better teamwork, and better teamwork with lower workload, but we found no evidence that battle-rostered crews exhibited better teamwork than mixed crews, and their average workload ratings were actually slightly higher than those of the mixed crews. We have not yet analyzed differences in the communications patterns of the battle-rostered and mixed crews.

Limited evidence on the effects of crew-coordination training from the pilot experiment shows that coordination training increased the mutual confidence of the crew members as well as increasing their perception that they could anticipate each other's decisions and actions and act in sync. We also found that coordination training improved ratings on many of the teamwork measures in the IP Post-Mission Teamwork Rating Form. Previous analysis (Entin, Entin, MacMillan, & Serfaty, 1993) has shown that

crew-coordination training can change communication patterns, leading to more anticipatory communication. Crew-coordination training has also been shown to improve crew performance (Simon & Grubb, 1993).

Plans for Future Analysis

The analysis reported above does not include the communication measures being derived from videotapes of the crews during the Battle-Rostering Experiment. Communication measures will be obtained for the time segments immediately before, during, and after the time segments for which workload data and detailed performance data are available. We will examine the volume and type of communications associated with higher and lower levels of teamwork, and the relationship between communication patterns and perceived workload. We will look for evidence that crews relied on implicit-coordination mechanisms such as anticipatory communication during high-task-demand periods, with a possible decrease in more-routine communications such as acknowledgments, and whether crews returned to more-explicit coordination patterns when task demand decreased. We will also examine links between the use of explicit and implicit coordination and indicators of the congruence of the crew's mental models such as the extent to which crew members felt they could anticipate each other's actions and the extent to which they felt they acted in sync. The communications analysis will provide insight into the mechanisms by which the high-teamwork crews were able to allocate their effort more effectively in order to achieve better performance in high-demand situations such as battle position operations.

References

Baker, D.P. & Salas, E. (1992). Principles for measuring teamwork skills. Human Factors, 34 (4), 469-475.

Cannon-Bowers, J.A., Salas, E., & Converse, S. (1990). Cognitive psychology and team training: Training shared mental models of complex systems. Human Factors Society Bulletin, 33 (12), 1-4.

Entin, E.B., Entin, E.E, MacMillan, J., & Serfaty, D. (1993). Structuring and training high-reliability teams: Year 1 Technical Report. Burlington, MA: ALPHATECH, Inc.

Grubb, G., Simon, R., Leedom, D. & Zeller, J. (1994). Evaluation of crew composition effects on AH-64 attack helicopter mission performance and flight safety. Wilmington, MA: Dynamics Research Corporation.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload. Amsterdam: Elsevier.

Kleinman, D.L. & Serfaty, D. (1989). Team performance assessment in distributed decisionmaking. In Gibson et al. (Eds.), Proceedings of the Symposium on Interactive Networked Simulation for Training. Orlando, FL.

LaPorte, T.R. & Consolini, P.M. (1988). Working in practice but not in theory: Theoretical challenges of high reliability organizations. Annual Meeting of the American Political Science Association. Washington, DC.

MacIntyre, R.M., Morgan, B.B., Jr., Salas, E.S., & Glickman, A.S. (1988). Teamwork from team training: New evidence for the development of teamwork skills during operational training. Paper presented at the Interservice/Industry Training Systems Conference, Orlando, FL.

Orasanu, J.M. (1990). Shared mental models and crew decision making (CSL Report 46). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Orasanu, J.M. & Fischer, U. (1992). Team cognition in the cockpit: Linguistic control of shared problem solving. In Proceedings of the 14th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Pfeiffer, J. (1989, July). The secret of life at the limits: cogs become big wheels, Smithsonian.

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. In D.E. Broadbent, J. Reason, & A. Baddely (Eds.), Human factors in hazardous situations. Oxford: Clarendon Press.

Serfaty, D., Entin, E.E., & Volpe, K. (1993a). Adaptation to stress in team decision making and coordination. In Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society (pp. 1228-1232). Santa Monica, CA.

Serfaty, D., Entin, E.E., & Volpe, K.(1993b). Implicit coordination in command teams. In Proceedings of the 1993 JDL Command and Control Research Symposium. McLean, VA: SAIC.

Simon, R. (1991). Technical report: Results of the data analysis Army aircrew coordination measures testbed conducted Spring 1990. Wilmington, MA: Dynamics Research Corporation.

Simon, R. & Grubb, G. (1993). Validation of crew coordination training and evaluation methods for Army aviation (E-785U). Wilmington, MA: Dynamics Research Corporation.

Appendix A

Crew Member Pre-Mission and Post-Mission Questionnaires

CREW MEMBER PRE-MISSION QUESTIONNAIRE

CREW ID: ____________	CREW POSITION: Pilot / CPG
DATE/TIME: ____________	LAST 4 DIGITS OF SS#: ________

1. Briefly list up to three potential "show stoppers" that could compromise this mission.

   1)____

   2)____

   3)____

2. Briefly describe the two most important tasks/responsibilities for you while at the battle position (BP).

   1)____

   2)____

3. Briefly describe the two most important tasks/responsibilities for your fellow crew member while at the BP.

   1)____

   2)____

On the scales below circle a number that best describes your attitude or belief.

4. How much confidence do you place in the ability of your fellow crew member to accomplish his role in this mission?

1	2	3	4	5	6	7
Very Low			Moderate			Very High

Explain briefly:____

5. How much confidence do you think your fellow crew member places in your ability to accomplish your role in this mission?

1	2	3	4	5	6	7
Very Low			Moderate			Very High

Explain briefly:____

CREW MEMBER POST-MISSION QUESTIONNAIRE

CREW ID: ____________	CREW POSITION:Pilot / CPG
DATE/TIME: ____________	LAST 4 DIGITS OF SS#: ________

On the scales below circle a number that best describes your attitude or belief.

1. How much confidence did you have in your fellow crew member, as compared to flying with other aviators in this unit?

1	2	3	4	5	6	7

Much Less — About The Same — Much More

2. How much assistance did you provide your fellow crew member, as compared to flying with other aviators in this unit?

1	2	3	4	5	6	7

Much Less — About The Same — Much More

3. How much did you cross-monitor the actions of your fellow crew member, as compared to flying with other aviators in this unit?

1	2	3	4	5	6	7

Much Less — About The Same — Much More

4. To what extent were you able to anticipate (i.e., predict) the actions and decisions of your fellow crew member?

1	2	3	4	5	6	7

Rarely — Half The Time — All The Time

5a. What was the most critical episode of this mission? ______________________

____

b. During this critical episode to what extent were you thinking and acting "in sync" with your fellow crew member?

1	2	3	4	5	6	7

Not At All — Moderately — A Great Deal

c. How do you know that? ______________________

____

Put an "X" on each of the six scales below, at the point that matches your workload experience.

6. Please rate the workload for the mission you just completed on the scales below:

Mental Demand

Very Low Very High

Physical Demand

Very Low Very High

Temporal Demand (Time Pressure)

Very Low Very High

Performance

Perfect Failure

Effort

Very Low Very High

Frustration

Very Low Very High

Appendix B

IP Post-Mission Teamwork Rating Form

IP POST-MISSION TEAMWORK RATING FORM

CREW ID:________________________	IP:______________
DATE/TIME:______________________	SCENARIO: 1 2 3 4

INSTRUCTIONS

Circle a number on the scale accompanying the questions below so that it best describes the behavior of the crew you just observed. Try to rate the behavior of the crew on an absolute scale and not a relative scale that compares one crew to the next. To help you perform this absolute rating a brief description of the behavior you should observe for the highest rating on the scale and a brief description of the behavior you should observe for the lowest rating on the scale are provided for each question. Read these guides or anchors carefully and refer to them as you rate the crew on each question. Feel free to write comments or explanations to any question.

Team Orientation

1. To what extent was this crew oriented toward teamwork?

1	2	3	4	5	6	7

7 Good team orientation could be inferred in a situation where a team member places the goals and interests of the team ahead of personal goals. Also may be evident through the display of trust, team pride, and esprit de corps, and an awareness that teamwork is important.

1 Poor team orientation manifests itself when members place their personal concerns above the team's success (e.g., disregarding or refusing to follow procedures; arguments, quarrels, and open resentment; and becoming upset with a member's performance and either ignoring or harassing that member are evidences of poor team orientation).

2. To what extent were errors caused by inadequate crew communication?

1	2	3	4	5	6	7

7 Communication within the crew was always effective and never responsible for errors or degraded performance.

1 Communication was wholly inadequate and resulted in most of the errors made by the crew.

3. To what extent were errors caused by inadequate individual actions?

1	2	3	4	5	6	7

7 No actions of a single crew member resulted in errors or poor crew performance

1 The actions and decisions by a single crew member very frequently resulted in errors or poor crew performance.

Comments:____

Communication Behavior

4. How clearly and timely did crew members communicate with each other?

1	2	3	4	5	6	7

7 Good communication occurs when team members pass on all important information and clarify intentions and planned procedures; communications are always clear, precise, and timely; members follow proper security procedures for communication; members always use proper terminology.

1 Poor communication occurs when crew members fail to pass on information or intentions, or pass on incomplete communications; members fail to clarify information; members disregard proper security procedures for communication; members use improper terminology; members tie up the net with irrelevant communications.

5. How well did crew members acknowledge each other's messages?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

7 Good acknowledgment occurs when team members acknowledge most if not all the messages sent to them; members obtain necessary information and acknowledge and repeat messages to ensure correctness; members ensure that their messages are received as intended.

1 Poor acknowledgment occurs when team members acknowledge few or none of the messages sent to them; members fail to acknowledge other member's requests or reports or fail to do it properly.

6. To what extent did the pilot provide relevant information to the CPG, without the CPG having to ask for it?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

7 Pilot always provided important information to the CPG without being asked.

1 Pilot never provided information to the CPG unless specifically asked.

7. To what extent did the CPG provide relevant information to the pilot, without the pilot having to ask for it?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

7 CPG always provided important information to the pilot without being asked.

1 CPG never provided information to the pilot unless specifically asked.

Comments:____

Monitoring Behavior

8. To what extent did crew members monitor each other's behavior?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

7 Good monitoring occurs when crew members consistently observe the performance of the other to ensure the efficiency of the team; members notice and are concerned with the performance of the entire team; one member recognizes when other crew member performs correctly; consistently keeps track of other crew member's performance.

1 Poor monitoring occurs when one crew member fails to notice the other's performance on almost all occasions; rarely notices when the other crew member performs correctly or makes a mistake.

9. To what extent did crew members alert each other to impending decisions and actions?

1	2	3	4	5	6	7

7 Crew members always alerted each other to impending decisions and actions; supporting information was actively solicited from other crew member.

1 Crew members did not keep each other informed of impending decisions and actions; compromises to flight safety or mission effectiveness arose when a crew member waited for the other to volunteer significant information.

Comments:____

Feedback Behavior

10. To what extent did crew members provide feedback to each other?

1	2	3	4	5	6	7

7 Good feedback behavior occurs when crew members go over procedures with each other by identifying mistakes and how to correct them; ask for input regarding mistakes and what needs to be worked on; members are corrected for mistakes and incorporate the suggestions in their procedures.

1 Poor feedback behavior occurs when one crew member makes sarcastic comments to other when the scenario doesn't go as planned; resists asking for advice and makes guesses on proper procedures; rejects time-saving suggestions offered by other crew member.

Comments:____

Backup Behavior

11. To what extent did crew members provide backup to each other?

1	2	3	4	5	6	7

7 Good backup behavior occurs when one crew member is having difficulty, makes a mistake, or is unable to perform duties, and the other member steps in to help ensuring that the activity is completed properly; one member provides critical assistance without neglecting their own assigned duties; member having difficulty or is overburdened displays a willingness to seek assistance rather than struggle and make a mistake.

1 Poor backup behavior occurs when one crew member fails to provide assistance to other member who is having difficulty, makes a mistake, or is unable to perform his duties; while providing assistance, the member tends to neglect his own duties; member is unwilling to ask for help even when it is available; one member provides needed assistance, but does not inform other that they are assisting of what they he has done; one member displays an unwillingness to help other even when asked.

12. To what extent did the pilot anticipate the need to provide task assistance to the CPG?

1	2	3	4	5	6	7

7 Pilot consistently anticipated the need to provide task assistance to CPG during critical phases of flight.

1 Pilot never anticipated the need to provide task assistance to CPG during critical phases of flight; the CPG always had to ask.

13. To what extend did the CPG anticipate the need to provide task assistance to the pilot?

1	2	3	4	5	6	7

7 CPG consistently anticipated the need to provide task assistance to pilot during critical phases of flight.

1 CPG never anticipated the need to provide task assistance to pilot during critical phases of flight; the pilot always had to ask.

14. Did the crew members adjust individual task responsibilities to prevent overload?

1	2	3	4	5	6	7

7 Crew members were consistently aware of workload buildup on each others and reacted quickly to adjust division of task responsibilities to redistribute workload among each other.

1 Crew members were generally unaware of workload buildup on each others; little or no attempt was made to adjust the distribution of task responsibilities before significant compromises to flight safety or mission effectiveness occur.

Comments:____

Coordination Behavior

15 To what extent was the crew's behavior coordinated?

1	2	3	4	5	6	7

7 Good coordination behavior occurs when one team member consistently passes critical information to the other member, thereby enabling him/her to accomplish tasks; one member consistently carries out tasks quickly or in a timely manner enabling other to carry out his tasks effectively. Crew members appear very familiar with the relevant parts of each other's job and carry out individual tasks in a synchronized manner.

1 Poor coordination behavior occurs when one team member consistently carries out his tasks ineffectively, leading to other team member failing at his tasks; members carry out their tasks unpredictably, leading to delays in execution of critical tasks; members neglect to pass on critical pieces of information to each other, leading to breakdown in team performance; team members carry out their tasks with significant delays leading to crew errors.

16. How congruent/similar were the pilot's and CPG's understandings of the mission?

1	2	3	4	5	6	7

7 Pilot and CPG were completely in agreement (i.e., congruent) on all goals, tasks, and concepts involving the mission.

1 Pilot and CPG were rarely in agreement (i.e., congruent) on all goals, tasks, and concepts involving the mission.

Comments:____

Appendix C

IP Performance-Rating Instrument by Time Segment

SCENARIO #1

IP EVALUATION OF PERFORMANCE FOR FIVE SITUATIONS

Make your ratings of perfomance using the same scale descriptors as you used for the ATM task ratings.

1-3. Recall the minor malfunction (#1 TRU Hot) that occurred in segment #2. Rate the crew's performance:

1. For approximately the two minute period just prior to ther onset of the minor malfunction.

   U S- S S+

2. During the minor malfunction emergency.

   U S- S S+

3. For approximately the two minute period just after the crew completed dealing with the minor malfunction.

   U S- S S+

4-5. Recall the crew's approach to the first Battle Position that occurred in segment #3. Rate the crew's performance:

4. For approximately the two minute period just prior to occupying the Battle Position.

   U S- S S+

5. During the Battle Position operation.

   U S- S S+

SCENARIO #2

IP EVALUATION OF PERFORMANCE FOR FIVE SITUATIONS

Make your ratings of perfomance using the same scale descriptors as you used for the ATM task ratings.

1-2. Recall the crew's approach to the first Battle Position that occurred in segment #3. Rate the crew's performance:

1. For approximately the two minute period just prior to occupying the Battle Position.

   U S- S S+

2. During the Battle Position operation.

   U S- S S+

3-5. Recall the major malfunction (#1 Loss of Oil) that occurred in segment #5. Rate the crew's performance:

3. For approximately the two minute period just prior to ther onset of the major malfunction.

   U S- S S+

4. During the major malfunction emergency.

   U S- S S+

5. For approximately the two minute period just after the crew completed dealing with the major malfunction.

   U S- S S+

SCENARIO #3

IP EVALUATION OF PERFORMANCE FOR FIVE SITUATIONS

Make your ratings of perfomance using the same scale descriptors as you used for the ATM task ratings.

1-3. Recall the minor malfunction (Util Hyd Oil Low) that occurred in segment #2. Rate the crew's performance:

1. For approximately the two minute period just prior to ther onset of the minor malfunction.

   U S- S S+

2. During the minor malfunction emergency.

   U S- S S+

3. For approximately the two minute period just after the crew completed dealing with the minor malfunction.

   U S- S S+

4-5. Recall the crew's approach to the first Battle Position that occurred in segment #3. Rate the crew's performance:

4. For approximately the two minute period just prior to occupying the Battle Position.

   U S- S S+

5. During the Battle Position operation.

   U S- S S+

SCENARIO #4

IP EVALUATION OF PERFORMANCE FOR FIVE SITUATIONS

Make your ratings of perfomance using the same scale descriptors as you used for the ATM task ratings.

1-2. Recall the crew's approach to the first Battle Position that occurred in segment #2. Rate the crew's performance:

1. For approximately the two minute period just prior to occupying the Battle Position.

   U S- S S+

2. During the Battle Position operation.

   U S- S S+

3-5. Recall the major malfunction (#1 CHIPs followed by engine failure) that occurred in segment #4. Rate the crew's performance:

3. For approximately the two minute period just prior to ther onset of the major malfunction.

   U S- S S+

4. During the major malfunction emergency.

   U S- S S+

5. For approximately the two minute period just after the crew completed dealing with the major malfunction.

   U S- S S+